An analysis of the use of digital collections in a scientific research library network: Part two of a case study from CSIC, Madrid, Spain

The introduction of the electronic journals in CSIC has been gradual and constant. There have been stable collections since 2002 and the data which is obtained from analysis of use has meaning especially in relation to the period 2002-2003. Developments in these years are presented below.

The proceedings for contracting the digital collection have been set down earlier but even though the objectives were clear it was also clear that it was necessary to count on the backing from the researchers as well as from the economic and political authorities of the institution. This conviction has meant that trial periods for each one of the platforms and statistics on use which guarantee interest for what is to be contracted always go along with any initiative related to the implantation of electronic journals.

IDEAL from Academic Press was the first platform of electronic journals which was made known to the CSIC community in a trial which took place in October-November 2000. The page which provided access to the trial from the Library Network Server included a brief questionnaire on opinions. This first survey showed that the users were familiar with electronic information resources and that they took a favourable view of electronic versions gradually replacing the paper collection in their centres. And it also showed attitudes and "desiderata" which have been maintained throughout these years:

- what is on offer is fine but they always suggest new titles and publishers
- the librarians are always more conservative with paper than the users and we/they want to guarantee physical copies for the future.
- the users perceived the global change brought by these electronic journals both in their work and in the library environment: it would be necessary to invest in computer equipment, it would decrease the demand for interlibrary lending, small research centres would benefit the most from the centralized acquisition of packets for everyone. Furthermore, four years ago they demanded an integrated, transparent presentation of the digital collection.

In 2002, ScienceDirect (Elsevier) and SpringerLink were contracted as well as IDEAL. There were free trials from Kluwer Online and WileyInterscience for two-three months to make these platforms known and these were each accompanied by evaluation sheets. The results of these surveys confirmed the interest in these products.

Electronic journals have advantages over printed ones when it comes to finding out the real use which is made of each title. Those responsible for the economy of the institution, without being specialists in research or librarianship, need to understand that the budget investment will bring yields. This is justified if we, the librarians are capable of demonstrating that periodic publications, which are so expensive, are used. On paper it has only been possible to respond to this interest in an intuitive way, approximately. Classic studies for evaluating use of the journal collections proposed very unreliable methods which did not show a real calculation at 100%. This obstacle has been overcome with electronic journals and we have accurate and contrastable information on which journals are used and which are not, what the core of the collection is and what can be dispensed with.

Statistics on electronic journals come from the servers of the publishers themselves, who count the accesses, the number of articles downloaded from each title, if indexes or summaries are consulted, if the texts are displayed in html or in pdf, etc. Nowadays we know more about the use of journals and the habits of the readers than ever before. In the case of CSIC, in the way it has carried out the contracting for publishers packets, this knowledge has limited use: even though we may know the titles which are not used we are not going to stop subscribing to them as the final price does not change. However, for our experience, this study of statistical reports has let us draw some comforting conclusions, such as the fact that what we have been buying on paper is also what our researchers consult most often in the electronic version, which is to say, we have a very relevant collection something which, up to now we had expected but which we could not prove. Moreover, we can see that there is a large number of titles never subscribed to on paper which have been highly successful with researchers and this justifies the initial decision not to only acquire online versions of subscriptions which already exist.

At the end of 2002 we drew up statistics on use of the platforms which were then available for the 12 months of the year (SpringerLink, ScienceDirect and Academic Press, the latter had been taken over by Elsevier half-way through the year), as well as statistics from other products which were used over shorter periods (WileyInterscience and Kluwer Online, or PCI Full Text). We did the same at the close of 2003 so that it was possible to observe the developments in use of these platforms over two years. (There are other new products -AIP, APS, Blackwell Synergy, MUSE, Nature, IOP- contracted for 2004. A first approximation of the statistics, made in April 2004, shows that the line of those acquired in the previous years is continuing and that Blackwell Synergy is the group which is receiving the most use).

Characteristics of the statistics on use of the electronic journals

The first comment is to show our appreciation for the existence COUNTER, a project which seems to have been established between the main science publishers since 2004. When we drew up statistics on use for the first time in 2002, the lack of uniformity in the measurement parameters between the different publishers in formal aspects as well as in the absence of standardization of terminology, made it extremely difficult to reach conclusions on the use which was being made of them. The only constant element was the *"articles downloaded as complete texts"*. This data was provided by all the publishers and it was the one taken as average in this study. When the editors provided more information (relation session/articles downloaded, localization by browse or by direct search, number of personal profiles registered, etc), this was also processed but it has to be accepted that few publishers got down to this type of detail in 2002.

For the CSIC –due to its structure and the form in which we have been contracting electronic journals- there are some short comings in the statistical reports usually provided by the publishers which are perhaps not as important for other institutions. Moreover, those of us working with this statistical information are experts in an office coordinating libraries, we are not in contact with the end users and it is impossible for us to have first-hand knowledge about the more than 4000 electronic journals we work with. In our opinion these reports show the following deficiencies:

- it is not usual to provided data by centre, but global use. The CSIC centres require individual information which we cannot give to them. When the publishers do in fact give this data by centre, it is not usually very definite as they mix our data with that of the universities and interpret accesses coming from the authentication systems erroneously.
- lists of journals which are not always those the CSIC subscribes to. On occasions the publishers offer their complete list to note the number of downloads on each line. But if a title had "0" downloads it may be that it is not an institution subscription and is not accessible for our users; it does not mean that they are not interested in it. Also, on certain occasions the lists which are sent by month or three-monthly periods vary in the number of journals they are giving information on, which makes it very difficult to match up the data title to title for the whole year.

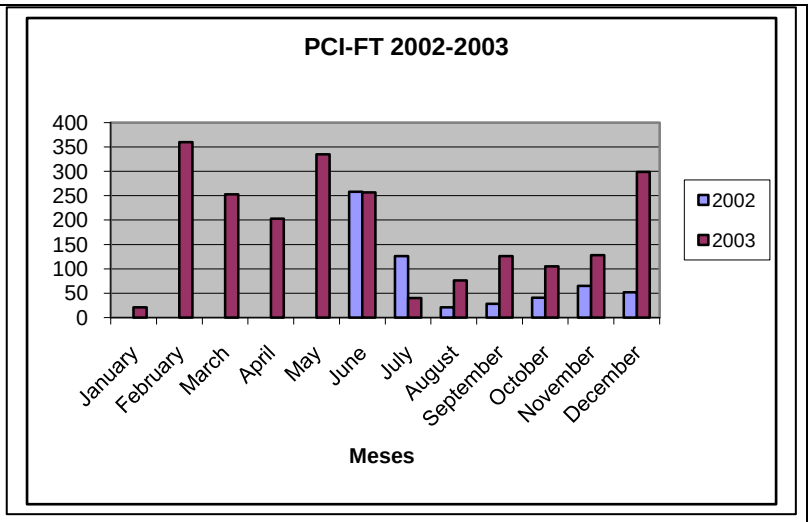
- little detail on rejected accesses. We would like to know which journals that are not subscribed to are required by the users or which years (retrospective) are requested but we do not have. When a complete packet has not been contracted this data is very useful and we miss having it.

- it would also be appreciated if these lists of downloaded titles/ articles included the data on coverage of the journal in such a way that it would be easy to know if it was a little used title because it closed some time ago and so its interest for the reader has diminished or if it has split or continued through others which do in fact register use.

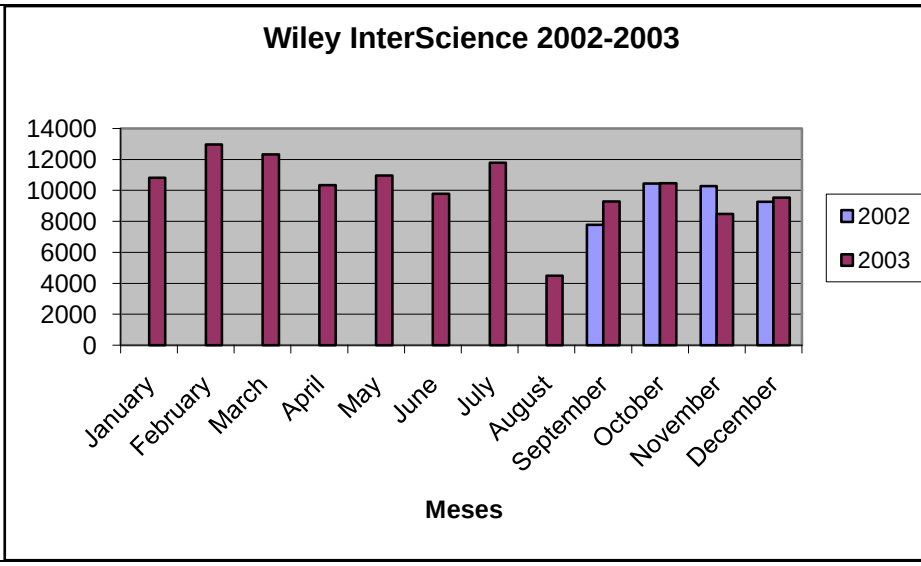
- and to continue making requests, vis-à-vis bibliometric studies, on the validity or obsolescence of the publications and on the interest of the retrospective archives, it would also be very useful for publishers to give information on the years for which consultation are made for each title.

Quantitative Data on use: articles downloaded in 2002/2003 in the CSIC by publishers:

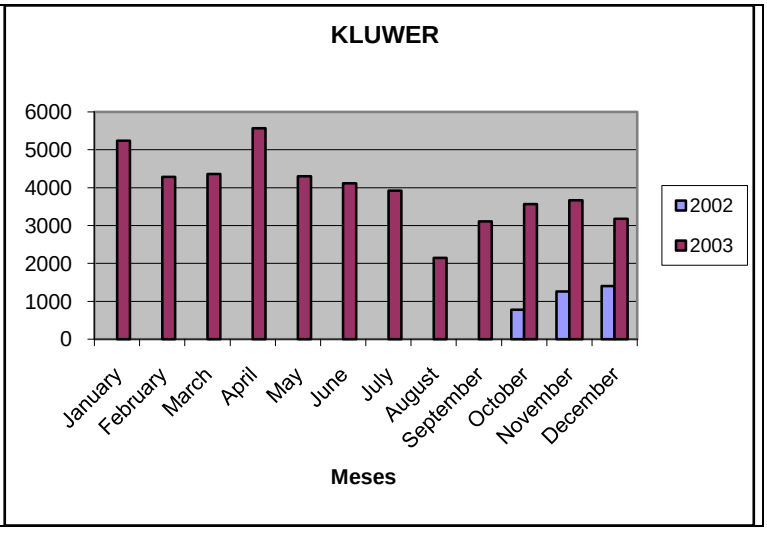
PCI-FT	2002	2003
January		21
February		360
March		253
April		203
May		335
June	258	257
July	126	40
August	21	76
September	28	126
October	41	105
November	65	128
December	52	299
Total	**591**	**2203**



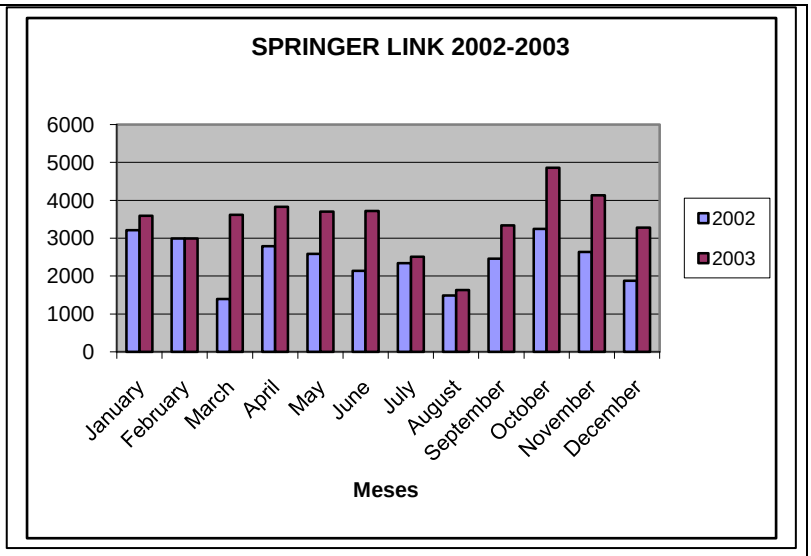
WILEY	2002	2003
January		10806
February		12975
March		12315
April		10350
May		10959
June		9782
July		11778
August		4482
September	7.781	9276
October	10.436	10467
November	10.269	8476
December	9.273	9537
Total	**37.759**	**121.203**



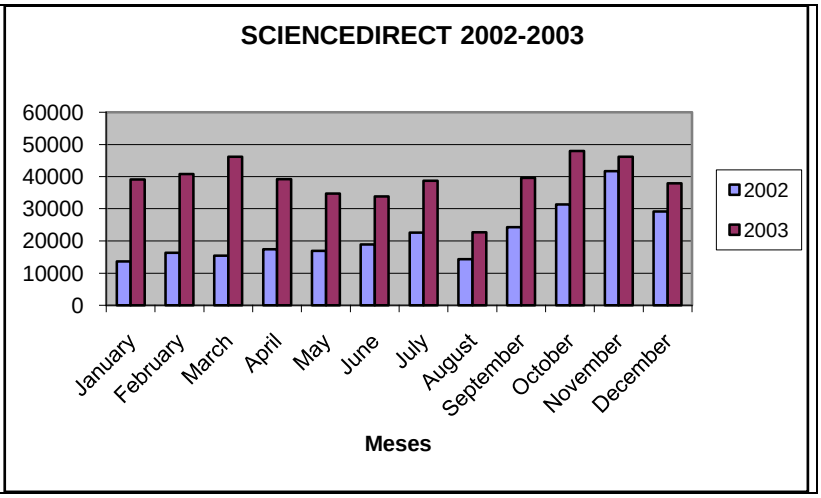
KLUWER	2002	2003
January		5240
February		4288
March		4362
April		5568
May		4299
June		4114
July		3920
August		2147
September		3110
October	775	3569
November	1263	3670
December	1.403	3178
Total	**3441**	**47465**



SPRINGER	2002	2003
January	3210	3592
February	2994	2994
March	1394	3617
April	2785	3829
May	2590	3705
June	2141	3718
July	2341	2507
August	1485	1628
September	2461	3339
October	3249	4855
November	2636	4130
December	1877	3276
Total	**29163**	**41190**



ELSEVIER	2002	2003
January	13662	39148
february	16364	40779
March	15449	46171
April	17407	39201
May	16936	34693
June	18878	33880
July	22617	38726
August	14326	22726
September	24284	39645
October	31347	47980
November	41696	46163
December	29213	37892
Total	**261999**	**467004**



The first three publishers were not available for the whole of 2002, but it is possible to see how they have been growing month by month, nearly doubling use for all of them (except Wiley) in the same month from one year to the next. The most striking data corresponds to ScienceDirect, which has to be set in relation to the quantity of journals there is in this platform and also, obviously its quality and the interest it arouses among CSIC researchers. An average of 270 articles was downloaded from ScienceDirect in 2003. (467,004 / 1728 journals); from Wiley (121,203 / 417 journals) 290, Kluwer (47,465 / 315) 150, Springer (41,190 / 429) 96 and PCI FT (2203 / 300) 7. The journal which was most successful among CSIC researchers was ***Angewandte Chemie International Edition***, from Wiley (11,769 downloads in 2003), followed by ***FEBS Letters***, from Elsevier (7,881 downloads in 2003). These were also the journals which were consulted most often in 2002. It is important to remember that chemistry is one of the most important research areas in the CSIC.

Searching habits

We consider that the module of use reports from Elsevier provides a greater quantity of information for the client. For this reason and because this platform is most used in the CSIC, we have drawn up a profile of the use habits of CSIC researchers from the reports obtained from this module of ScienceDirect:

(a) – ***average time*** *used* ***in searches:*** this was becoming shorter and shorter which shows that the users were becoming more skilful (in January, 2002 the average was 14 minutes; in January 2003 it was 7:11 minutes, and in January 2004 it remained at 7:24 minutes)

(b) – the users ***navigate very much in the pages*** (in 2002 1,437,701 pages visited ; in 2003 2,168,546), until the articles to be downloaded were found (262,089 articles in 2002 and 467,004 in 2003)

(c) - the **preferred format** for downloading the articles is PDF; browse was the **type of search** used most frequently. The articles found by direct search were fewer than those found with browse.

Search / Article	2002	2003
Art. / browse	236938	407572
Art./ search	25061	59432
PDF	221849	352320
HTML	40150	114684
Abstracts	26038	45163

The direct search can be of different types: quick, basic, or advanced. The one which the users employ most often is the "rapid" (92,151 in 2003), followed by the basic (89,160) while the advanced which is used by boolean operators hardly has any success (4,913 occasions).

(d) – Searches in **other types of content**: Requests for SummaryPlus, Abstract & References and Abstracts (no full text articles). In this case nearly all the journals have received some access. In 2003 there was a total of 88,915 consultations of abstracts, summaries, etc. and only 13 journals had "0" consultations.

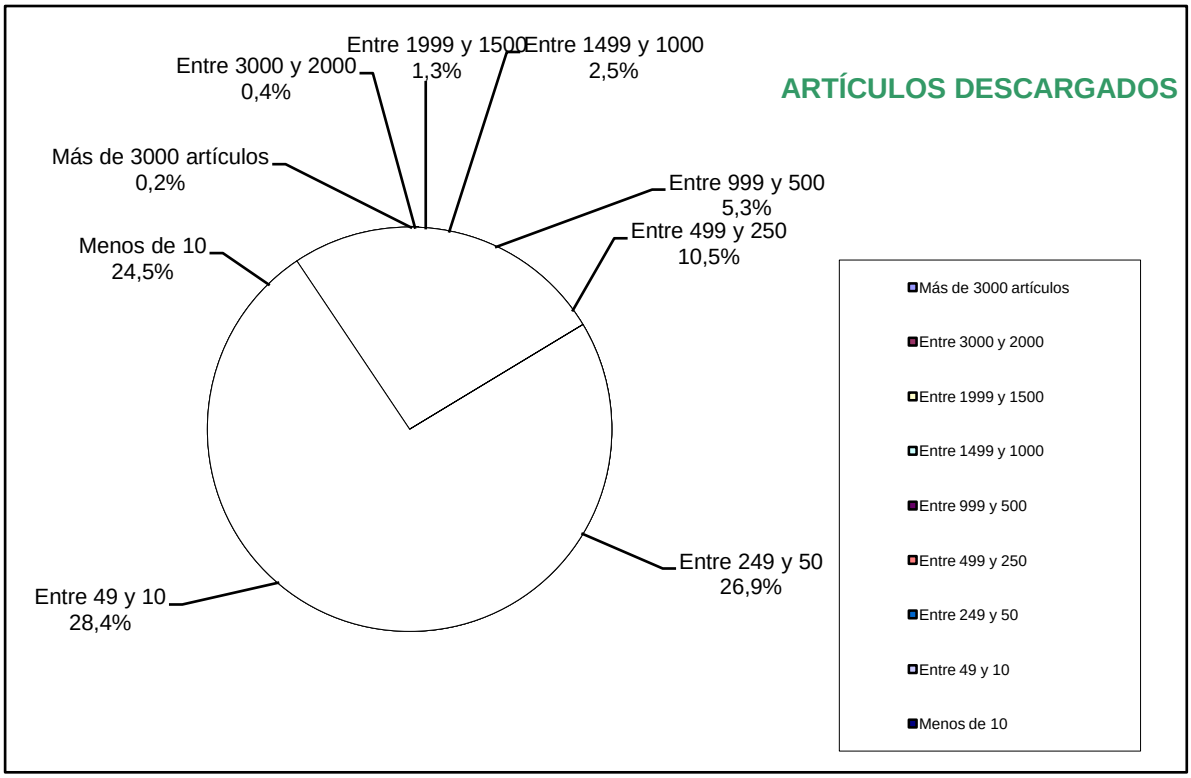
It is very interesting that some of the journals which are in the leading positions by number of complete articles downloaded hardly had any abstracts consulted and the reverse, there are titles which are in important positions in the consultation of summaries or abstracts but do not have the same position when it comes to ordering by articles downloaded.

Analysis of use of the collection

The ScienceDirect statistics module (now also from many publishers in compliance with COUNTER) offers an alphabetic list of titles available with the number of articles which have been downloaded throughout the year. We have organized this list by quantity of articles downloaded, so obtaining a series of data which is shown in the following tables and graphs where we compare the developments in use in the years 2002 and 2003.

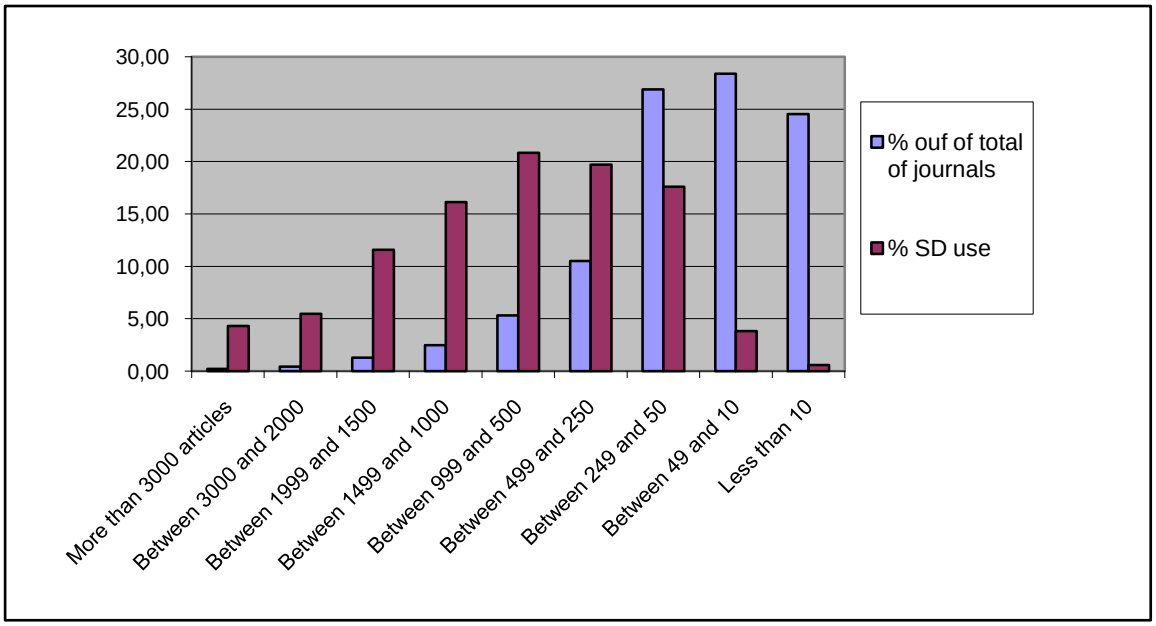
2002

Range of articles	Journals	% out of total of journals	No. of articles per range	% use of ScienceDirect
More than 3000 articles	3	0.21	11294	4.31
Between 3000 and 2000	6	0.43	14370	5.48
Between 1999 and 1500	18	1.28	30346	11.58
Between 1499 and 1000	35	2.48	42272	16.13
Between 999 and 500	75	5.32	54613	20.84
Between 499 and 250	148	10.50	51594	19.69
Between 249 and 50	379	26.87	46094	17.59
Between 49 and 10	400	28.37	9987	3.81
Fewer than10	346	24.54	1429	0.57
TOTAL	1410	100.0	261999	100.0



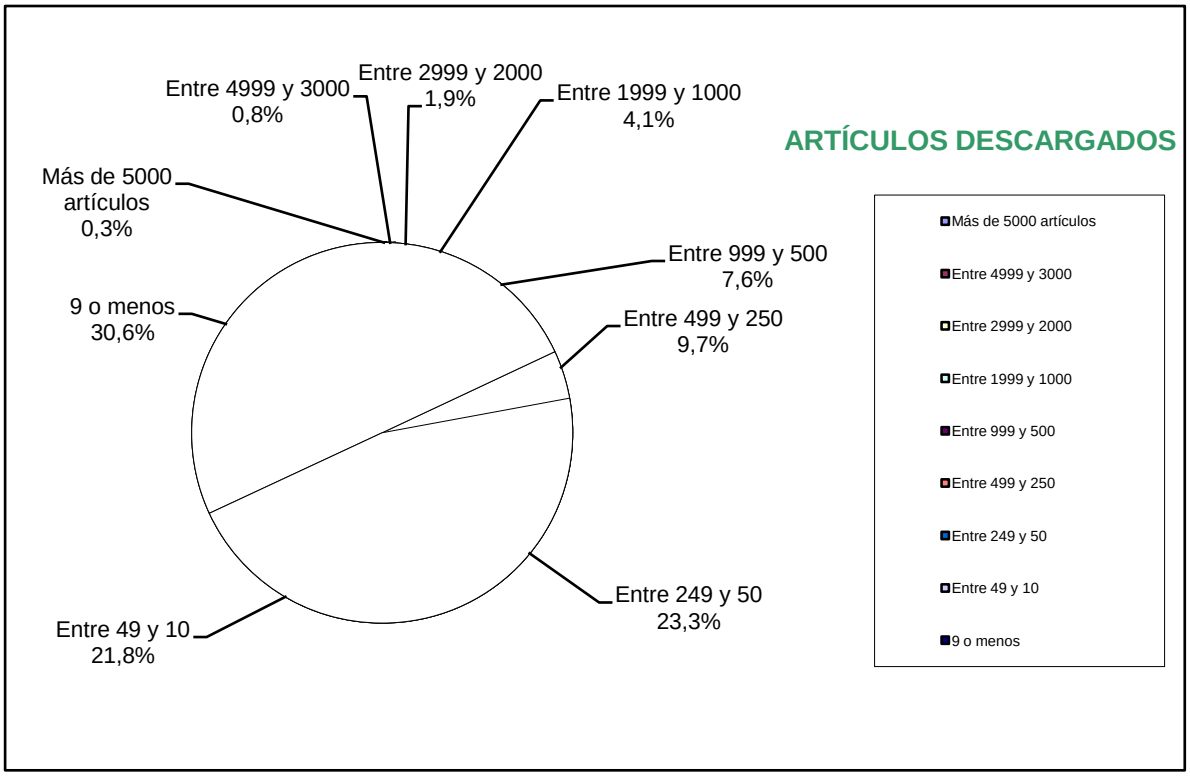
If we look at a graph with only these two parameters (percentage of journals and percentage of articles out of the total of those downloaded), we can see how the columns in the two colours rise and fall in opposite directions:

	% out of total of journals	**% SD use**
More than 3000 articles	0.21	4.31
Between 3000 and 2000	0.43	5.48
Between 1999 and 1500	1.28	11.58
Between 1499 and 1000	2.48	16.13
Between 999 and 500	5.32	20.84
Between 499 and 250	10.50	19.69
Between 249 and 50	26.87	17.59
Between 49 and 10	28.37	3.81
9 or fewer	24.54	0.57
Total	100.0	100.00

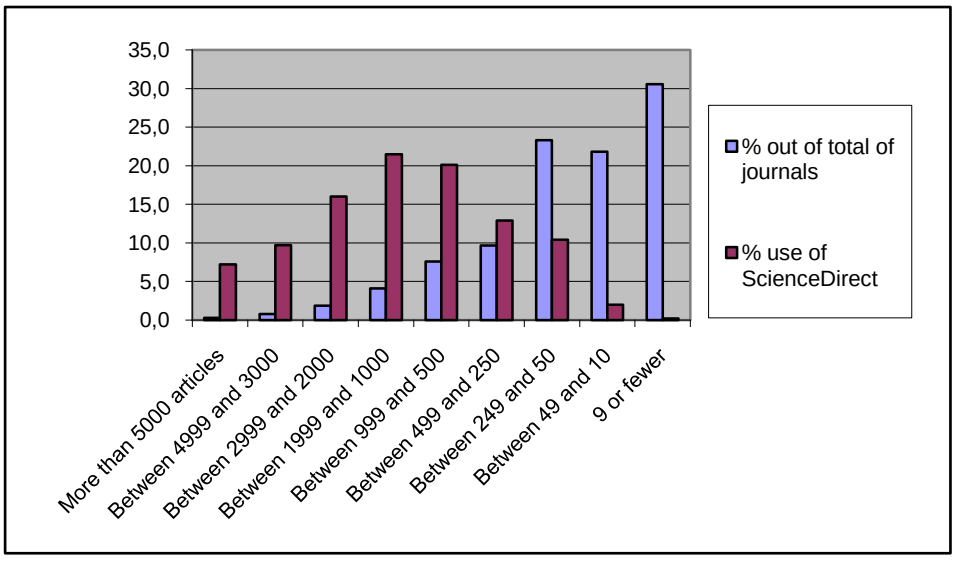


2003

Range of articles	Journals	% out of total of journals	No. of articles per range	% use of ScienceDirect
More than 5000 articles	5	0.3	32992	7.1
Between 4999 and 3000	13	0.8	45329	9.7
Between 2999 and 2000	32	1.9	74832	16.0
Between 1999 and 1000	71	4.1	100599	21.5
Between 999 and 500	131	7.6	93886	20.1
Between 499 and 250	167	9.7	60129	12.9
Between 249 and 50	403	23.3	48779	10.4
Between 49 and 10	377	21.8	9387	2.0
9 or fewer	529	30.6	1072	0.2
TOTAL	1728	100.0	467005	100.0



	% out of total of journals	% use of ScienceDirect
More than 5000 articles	0.3	7.2
Between 4999 and 3000	0.8	9.7
Between 2999 and 2000	1.9	16.0
Between 1999 and 1000	4.1	21.5
Between 999 and 500	7.6	20.1
Between 499 and 250	9.7	12.9
Between 249 and 50	23.3	10.4
Between 49 and 10	21.8	2.0
9 or fewer	30.6	0.2
Total	100.0	100.0



The first figures in these tables show that in both 2002 and 2003 they complied with a traditional enunciation in the evaluation of collections, which had been drawn up for journals on paper, but which is also repeated for the electronic collections: 80% of use is concentrated on 20%.

In 2002, the first six ranges of journals (up to the parameter 499-250 articles downloaded) represent 78.03% of use out of 20.22 % of the journals available. In 2003 these same 6 ranges were 87.40 % of use out of 24.4 % of the total of journals.

There is clearly a core of about 50 journals in this ScienceDirect service which are essential for CSIC researchers, titles which repeat their position in the internal journal ranking which we have drawn up over the two years studied.

We also see that during these two years there are still 50% of the titles (those from which less than 49 articles have been downloaded) which we could say have little or not interest for CSIC researchers. This is a very high percentage which shows that it is necessary to reconsider the titles which are being acquired. However, so as not to be so negative, let us remember that abstracts or summaries have been consulted from all the journals, except 13, as stated earlier, although they have not been interesting enough to download a complete article. We have also realized that the titles which show ”0” use (291 in 2003) are subject to various situations:

- There are several titles which are part of a packet from 5 or 10 journals, each of them on one subject, in which the speciality, which is not present in the CSIC, is not used.
- General titles, with other more specific articles under this umbrella and these are used (possibly the general titles has been split into part A and part B which are consulted.)
- cases of journals which are not subscribed to by the institution
- supplements, letters or addenda which have the same titles as other journal which are frequently consulted but whose “letters” or “reports” do not have any interest

Whatever the case these are journals which are paid for but not used.

Comparison with subscriptions on paper

The CSIC subscribes in a centralized way, through the procedure of public tenders, 415 titles on paper from Elsevier (the number of subscriptions is higher due to the duplicates). The comparison between the journals which are bought on paper and this internal ranking of the most frequently used titles illustrates the relevance or not of the titles which the CSIC has been subscribing to for years in the printed version.

Range of articles	**Journals**	**Subscribed on paper**	**Not subscribed on paper**	**% subscribed on paper**
More than 5000 articles	5	5	0	100 %
Between 4999 and 3000	13	8	5	61.53 %
Between 2999 and 2000	32	25	7	78.12 %
Between 1999 and 1000	71	57	14	80.28 %
Between 999 and 500	131	81	50	61.83 %
Between 499 and 250	167	76	91	45.50 %
Between 249 and 50	403	104	299	25.80 %
Between 49 and 10	377	39	338	10.34 %
9 or fewer	529	20	509	3.78 %
TOTAL	**1728**	**415**	**1313**	

If we remain at a ratio of 80% of use out of 20% of the collection which has served our purposes beforehand we see that in 2003, 87.40 % of use (24.4 % of the collection) is borne by 419 journals. Of these CSIC subscribes to 252 on paper and does not subscribe to 167. This means that the CSIC has subscribed, on paper, to 60.14 % of the journals that interest it most. And also, thanks to the acquisition of

ScienceDirect as a whole, researchers consult nearly 40% of the journals they did not have before. If we limit ourselves to the first ranges the percentage relevance of the collection on paper is much higher; there are 121 journals from which more than 1000 articles have been downloaded (54.40 % of the total use of ScienceDirect). Of these the CSIC subscribes to 115 on paper, that is to say, more than 95% of the titles which hold interest.

At the end of this list there are 529 journals (30.6 % of the total of ScienceDirect) with insignificant use, fewer than 9 articles have been downloaded. They represent 0.2 % of the demand on Elsevier. Out of these 346 journals, 20 are subscribed to on paper.

The drop in interlibrary loan in the CSIC Library Network has to be set in relation to the existence of all these available journals (not only in Elsevier, the percentages noted are similar in the other publishers) in electronic version which have not been in our libraries before.

The balance of the relevance of the collection on paper, in the light of this data, is quite positive, it could be concluded that for decades the CSIC has been subscribing to the titles from Elsevier which are most requested by its researchers. The next step should be to coordinate, in the most rational and economical way possible, the coexistence of this printed collection with the digital collection.

But it must be taken into account that two years is still a very short time to establish trends in use. However, there is one thing that is clearly shown and that is that electronic journals are perfectly implanted in the community of CSIC users.